Conformational plasticity of hepatitis C virus core protein enables RNA-induced formation of nucleocapsid-like particles

Holmstrom, Erik D ; Nettels, Daniel ; Schuler, Benjamin

Abstract: Many of the unanswered questions associated with hepatitis C virus assembly are related to the core protein (HCVcp), which forms an oligomeric nucleocapsid encompassing the viral genome. The structural properties of HCVcp have been difficult to quantify, at least in part because it is an intrinsically disordered protein. We have used single-molecule Förster Resonance Energy Transfer techniques to study the conformational dimensions and dynamics of the HCVcp nucleocapsid domain (HCVncd) at various stages during the RNA-induced formation of nucleocapsid-like particles. Our results indicate that HCVncd is a typical intrinsically disordered protein. When it forms small ribonucleoprotein complexes with various RNA hairpins from the 3' end of the HCV genome, it compacts but remains intrinsically disordered and conformationally dynamic. Above a critical RNA concentration, these ribonucleoprotein complexes rapidly and cooperatively assemble into large nucleocapsid-like particles, wherein the individual HCVncd subunits become substantially more extended.

DOI: https://doi.org/10.1016/j.jmb.2017.10.010

Conformational plasticity of hepatitis C virus core protein enables RNA-induced formation of nucleocapsid-like particles

Erik D. Holmstrom[a]*, Daniel Nettels[a], and Benjamin Schuler[a,b]*

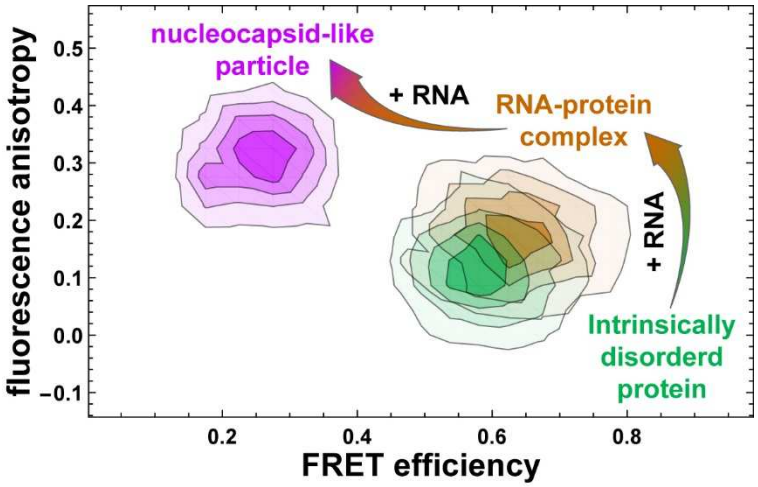
[a]Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
[b]Department of Physics, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
* Corresponding Authors

Erik Holmstrom	Tel: +41 44 635 5534;	Fax: +41 44 635 5907;	Email: e.holmstrom@bioc.uzh.ch
Daniel Nettels	Tel: +41 44 635 6553;	Fax: +41 44 635 5907;	Email: nettels@bioc.uzh.ch
Benjamin Schuler	Tel: +41 44 635 5535;	Fax: +41 44 635 5907;	Email: schuler@bioc.uzh.ch

ABSTRACT

Many of the unanswered questions associated with hepatitis C virus assembly are related to the core protein (HCVcp), which forms an oligomeric nucleocapsid encompassing the viral genome. The structural properties of HCVcp have been difficult to quantify, at least in part because it is an intrinsically disordered protein (IDP). We have used single-molecule FRET techniques to study the conformational dimensions and dynamics of the HCVcp nucleocapsid domain (HCVncd) at various stages during the RNA-induced formation of nucleocapsid-like particles. Our results indicate that HCVncd is a typical IDP. When it forms small ribonucleoprotein complexes with various RNA hairpins from the 3′ end of the HCV genome, it compacts but remains intrinsically disordered and conformationally dynamic. Above a critical RNA concentration, these ribonucleoprotein complexes rapidly and cooperatively assemble into large nucleocapsid-like particles, wherein the individual HCVncd subunits become substantially more extended.

GRAPHICAL ABSTRACT



HIGHLIGHTS

- What role does the core protein play in nucleocapsid-like particle (NLP) assembly?
- The core protein is highly dynamic and unstructured in the absence of RNA
- A small, dynamic ribonucleoprotein complex forms at low RNA concentrations
- At high RNA concentrations, these ribonucleoprotein complexes assemble into NLPs
- The core protein adopts an expanded structure within the oligomeric NLPs

1.0 INTRODUCTION

The hepatitis C virus (HCV) is a widespread pathogen that infects more than 1 in 50 people worldwide and is responsible for a range of progressive hepatotropic diseases **[1]**, including a quarter of all instances of liver cancer **[2]**. The virus particle is a 50-60 nm diameter, lipid-enveloped nucleocapsid **[3]** (Figure 1a) that encompasses the viral genome, which is a ~ 9600 nucleotide, positive-sense, single-stranded RNA **[4]** (Figure 1b). This RNA is directly translated by the host ribosome via an internal ribosomal entry site (IRES), resulting in a single ~ 3000 amino acid polyprotein that is cleaved by both host and viral proteases into a total of 10 viral proteins **[5]** (Figure 1c); three of which—core protein (HCVcp), envelope protein 1 (E1), and envelope protein 2 (E2)—are responsible for the structural integrity of the fully assembled viral particle (Figure 1b).

Viral particle assembly, in particular nucleocapsid formation, is driven by interactions between genomic RNA and the multifunctional HCVcp. This process represents a promising therapeutic target in the HCV life cycle that has not yet been fully exploited because many critical details remain unknown **[5]**. In particular, our understanding of nucleocapsid formation can greatly benefit from more structural information **[6,7]**. However, such experiments are often challenging for intrinsically disordered proteins (IDPs) like HCVcp **[8]**. Although IDPs are commonly found in viral proteomes **[9]**, where their conformational plasticity allows them to perform a diverse collection of biological functions **[10]**, their pronounced dynamic structural heterogeneity often makes them difficult to characterize using traditional ensemble techniques.

One successful experimental technique for characterizing IDPs is single-molecule Förster Resonance Energy Transfer (FRET), which can be used to measure the conformational dynamics and dimensions of these disordered systems **[11,12]**. Unlike techniques that average over the entire ensemble, single-molecule methods can identify and characterize distinct subpopulations that may be present at equilibrium. This is particularly important for IDPs, which can transiently populate multiple conformations. Here we use single-molecule FRET to perform a quantitative investigation of the conformational dynamics and dimensions associated with the core protein's nucleocapsid domain (HCVncd) at various points during RNA-induced formation of nucleocapsid-like particles (NLPs). In the absence of RNA, we find that HCVncd is largely unstructured and highly dynamic, making it a typical IDP. At low RNA concentrations, this IDP forms a small ribonucleoprotein complex with structured RNAs from the viral genome. Surprisingly, HCVncd appears to remain intrinsically

disordered and highly dynamic within this complex. At higher RNA concentrations, these complexes rapidly and cooperatively assembly into much larger NLPs, wherein the HCVncd subunits become substantially more expanded.

2.0 RESULTS AND DISCUSSION

2.1 HCVncd in the absence of RNA

The hepatitis C virus core protein (HCVcp) is comprised of the first ~ 191 amino acids of the host-translated polyprotein (Figure 1c). HCVcp is a highly conserved, positively charged protein that oligomerizes around genomic RNA to form the viral nucleocapsid **[13]**. In addition to its role as a structural protein, HCVcp is also a nucleic acid chaperone that promotes viral genome dimerization **[14-16]** and a biomolecular hub that is thought to mediate critical virus-host interactions **[17]**. It consists of two domains (Figure 1d). The N-terminal nucleocapsid domain of HCVcp (HCVncd) is highly conserved, positively charged, and intrinsically disordered (Supplemental Figure 1)**[8]**. This region of the protein is both necessary and sufficient for: (i) nucleocapsid-like particle (NLP) formation **[18]** and (ii) nucleic acid chaperone activity **[14-16]**. The more hydrophobic C-terminal domain is thought to direct HCVcp to lipid droplets **[19]** and the endoplasmic reticulum, the putative location of viral particle assembly **[20]**. Since the primary focus of this investigation centers around NLP assembly, we used site-specific mutagenesis to introduce two cysteines (S2C, T65C, C98M) into the N-terminal domain of a clinically isolated HCVncd variant (Figure 1e) **[21]**. These cysteines are then coupled to donor (Alexa 488) and acceptor (Alexa 594) fluorophores for single-molecule FRET experiments. In particular, these labelling sites allow us to sensitively probe the conformational dimensions and dynamics associated with this highly conserved region of HCVncd (Supplemental Figure 1) at various stages during NLP assembly.

Circular dichroism (CD) experiments (Figure 2a) were used to characterize the degree of disorder associated with HCVncd. In accordance with previous CD spectra acquired for slightly different segments of HCVcp **[8,15,22]**, HCVncd displays a minimum in ellipticity below 200 nm–a hallmark of polypeptides lacking secondary structure. A reconstruction of the CD spectrum **[23]** using protein reference sets **[24-26]** suggests that the helical content of this sequence is between 4% and 8%, similarly low as the previously studied constructs **[15,22]**.

To quantify the conformational dimensions and dynamics of HCVncd, we use diffusion-based single-molecule FRET experiments **[27-29]**. In phosphate-buffered saline, single FRET-labeled HCVncd molecules (~100 pM) transiently diffuse through the confocal volume, giving rise to ~1 ms long bursts each containing 50 - 100 fluorescence photons. The transfer efficiency, *E*, can be calculated ratiometrically using the number of donor and acceptor photons in each burst (see MATERIALS AND METHODS). The results from thousands of bursts yield an *E* histogram (Figure 2b, green; Supplemental Figure 2) with a single peak and a mean value of $\langle E \rangle = 0.62 \pm 0.01$. The addition of either GdmCl or NaCl (Figure 2b and Supplemental Figure 3, respectively) results in a continuous shift of the peak, as expected for an intrinsically disordered protein. As previously observed for other IDPs**[27,28,30]**, GdmCl shifts the *E* histograms to lower values (Figure 2b), corresponding to an expansion of the polypeptide chain **[31,32]**, whereas NaCl shifts the *E* histograms to slightly higher values (Supplemental Figure 3). This continuous shift of a single population in the *E* histograms is in stark contrast to the redistribution of discrete subpopulations (e.g., folded and unfolded) associated with the cooperative folding/unfolding of proteins with large free energy barriers and interconversion kinetics **[33]**.

Although the mean value of the *E* histograms can be quantitatively related to the average distance between the donor and acceptor fluorophores, the histogram width is dominated by shot-noise arising from the finite number of photons, because the intramolecular distance dynamics of unfolded and intrinsically disordered proteins are faster than the average interphoton time **[34,35]**. However, information about the width of the underlying FRET efficiency distribution (and thus the distance distribution) is contained within the fluorescence lifetimes of the fluorophores, because the distance dynamics in unstructured proteins are slower than the fluorescence lifetimes of fluorophores**[36-39]**. Lifetime information is available from the single-molecule data since they are recorded using time-correlated single-photon counting (TCSPC) **[40]** and pulsed interleaved excitation **[41]**. In two-dimensional density plots of normalized lifetimes vs transfer efficiency (Figure 2c), systems with a single fixed FRET efficiency are expected to lie on the diagonal black line in Figure 2c. However, when transfer efficiency fluctuations occur on a timescale between the donor fluorescence lifetime (~ 4 ns) and the inter-photon time (~ 10 μs), lifetimes for each event will cluster either above the diagonal if donor photons are analyzed or below the diagonal if acceptor photons are analyzed (Figure 2c, blue

and orange, respectively). The vertical displacement from the diagonal can be related quantitatively to the variance of the underlying FRET efficiency distribution, σ^2 (see MATERIALS AND METHODS).

An analysis of a lifetime density plot containing thousands of bursts (Figure 2c) reveals that the variance of the FRET efficiency distribution of HCVncd under near-physiological conditions is $\sigma^2 = 0.13 \pm 0.02$. As commonly seen in unfolded and intrinsically disordered proteins **[34]**, this variance **[38,39]** arises from conformational dynamics within each burst, which gives rise to a distribution of inter-dye distances, $P(r_{\text{dye-dye}})$. Simple polymer models provide a remarkably accurate approximation of these intramolecular distance distributions in unfolded and intrinsically disordered proteins **[34]**. This has been shown by experiments that probe multiple segments in the same protein **[31,42,43]**, comparisons between FRET-derived distributions and distributions obtained using other techniques, like NMR **[42]** and SAXS **[31,42]**, and distributions obtained using the most realistic all-atom molecular dynamics force fields currently available **[31,44,45]**. For this work, we chose to use a wormlike-chain (WLC) model to describe $P(r_{\text{dye-dye}})$, because it has a well-defined maximal value for $r_{\text{dye-dye}}$, which is useful for calculating the expected behavior in the lifetime density plots. For a contour length of l_c = 274 Å (corresponding to 64 amino acids plus the dyes), our experimentally determined transfer efficiency of $\langle E \rangle = 0.62 \pm 0.01$ corresponds to a persistence length of $l_p \approx 4.8$ Å, resulting in a $P(r_{\text{dye-dye}})$ with a mean value of $\langle r_{\text{dye-dye}} \rangle \approx 49$ Å (see MATERIALS AND METHODS). Most importantly, this $P(r_{\text{dye-dye}})$ gives rise to a FRET efficiency distribution with a variance of $\sigma^2 = 0.11$, which is consistent with the variance identified via the lifetime density plot*. This finding indicates that HCVncd molecules rapidly sample a broad intramolecular distance distribution during their transit through the confocal detection volume (Figure 2c). However, it does not necessarily imply that HCVncd corresponds to a fully random coil, because we only probe one large polypeptide segment (i.e., 64 amino acids), which will result in averaging over any residual transient structure that may exist within the chain.

In this way, the continuous shift of the E histograms resulting from the addition of GdmCl and NaCl can also be quantitatively analyzed to determine the corresponding change in $\langle r_{\text{dye-dye}} \rangle$, which illustrates the resulting expansion or compaction of the chain, respectively (Figure 2d). Interestingly, at low salt concentration, the dimensions of the FRET-labeled segment of HCVncd correspond to a

* The variance of the FRET efficiency distribution is also consistent with other simple polymer models (e.g., Gaussian chain, Self-avoiding Chain, etc.), because the shapes of the different distance distributions are very similar if l_c is much larger than l_p, as in the case of HCVncd.

length scaling exponent of $\nu \approx 0.49$ (see MATERIALS AND METHODS), very near the value of 1/2 expected for an ideal chain under theta conditions where attractive intra-chain interactions balance the repulsive excluded volume interactions**[28]**. As previously reported for other IDPs, the enhanced electrostatic screening caused by the addition of NaCl weakens the repulsive interactions between the many positively charged amino acids in HCVncd, allowing the polypeptide to more easily sample compact, and thus charge dense, conformations**[27]**. Denaturants, such as GdmCl, are thought to bind to and more effectively solubilize the polypeptide backbone, which gives rise to the well-document expansion of unfolded and unstructured proteins**[28,31,32]**.

The lifetime density plot (Figure 2c) indicates that HCVncd rapidly samples a distribution of intramolecular distances. This hypothesis can be further tested with nanosecond FCS techniques **[46]**, which are sensitive to fluctuations in the donor and acceptor photon detection rates resulting from chain reconfiguration. Indeed, the donor-donor ($g(\tau)_{DD}$) and acceptor-acceptor ($g(\tau)_{AA}$) cross correlation functions show pronounced correlated behavior over the course of 150 ns (Figure 2e, blue and orange), and the donor-acceptor cross correlation function ($g(\tau)_{DA}$) shows an anti-correlated component (Figure 2e, grey) on the same timescale, as expected for an IDP that samples a distribution of $r_{dye\text{-}dye}$ on a this timescale **[29,47]**. A global analysis of the nanosecond cross correlation functions results in a chain reconfiguration time of $\tau_r = 47 \pm 3$ ns, thus defining the timescale associated with the rapid conformational dynamics of HCVncd.

These biophysical experiments quantitatively characterize the dimensions and dynamics of HCVncd in the absence of RNA. They show that HCVncd does not have any persistent secondary structure, but instead that it samples a distribution of FRET efficiencies with the expected variance of an unstructured chain. As previously observed for other IDPs **[29,44]**, our results indicate that HCVncd samples such a distribution on a ~50 ns timescale. In the context of the well-established structure-function paradigm of molecular biology, it is remarkable that this unstructured protein can adopt multiple functional roles during the HCV life cycle **[17]**, including that of an RNA chaperone **[14-16]** as well as its namesake role as the primary constituent of HCV nucleocapsids **[13]**. We now turn to these functional interactions of HCVncd.

2.2 HCVncd in the presence of RNA

Formation of nucleocapsid-like particles (NLPs) can be induced *in vitro* with high concentrations of structured RNAs **[18]**. For this study, we use four small viral RNAs (RNA1, RNA2, RNA3, RNA4; Figure 3a) from the 99 nt untranslated 3′ X-RNA at the end of the viral genome (Figure 1b, black). We focused on this region because it is known to be part of an intricate RNA-RNA and RNA-protein interaction network that regulates various aspects of the viral life cycle **[48-51]**, including viral assembly **[52]**. All four RNAs contain the dimer linkage sequence (DLS; Figures 1b and 3a, cyan), which is a 16 nt self-complementary sequence that can form a homodimeric kissing complex **[53]**. This kissing complex, with the assistance of HCVcp acting as an RNA chaperone **[14,54]**, rearranges into an extended RNA duplex thus linking two copies of the genome. Notably, a dimeric form of the genome has been proposed to be an important component of the aforementioned regulatory network **[48,53,55]**.

To determine whether the conformational heterogeneity of HCVncd in isolation (designated as "FREE") persists throughout the NLP assembly process, increasing amounts of each nucleic acid were added to solutions of FRET-labeled HCVncd to achieve final RNA concentrations between 0 μM and 20 μM. In addition to the ~100 pM labeled HCVncd, all samples also contained 3 nM unlabeled HCVncd to minimize potential intermolecular FRET between the individual HCVncd subunits within the fully assembled NLPs (i.e., capsomeres). The *E* histograms for these four data sets (Figure 3b) show that HCVncd can sample at least three different populations during RNA-induced NLP assembly. The RNA-dependent stability of these three populations provides new structural insights into the conformational plasticity of HCVncd during RNA-induced NLP assembly.

2.2.1 Formation of RNA-HCVncd ribonucleoprotein complexes at low concentrations of RNA

Low concentrations (i.e., < 4 μM) of all four RNAs cause the FREE population (Figure 3, green) to transition to a population associated with small ribonucleoprotein (RNP) complexes (Figure 3, orange) at increased transfer efficiency, indicating a subtle RNA-induced compaction of HCVncd. The extent of compaction is so small for the three longest RNAs (i.e., RNA2, RNA3, and RNA4) that the FREE and RNP populations overlap substantially. Therefore, the RNA-induced redistribution of these two populations results in a single broadened peak that undergoes a small ($\Delta E \approx 0.05$) but significant shift (relative to our experimental precision of ~0.01) to higher transfer efficiencies. A global analysis of the *E* histograms enables us to quantify the mean transfer efficiency of the RNP population, $\langle E \rangle_{RNP}$, as

well as its fractional abundance, f_{RNP}, as a function of RNA concentration (see MATERIALS AND METHODS). Moreover, the transition from FREE to RNP can independently be monitored by an increase in the fluorescence anisotropy of labeled HCVncd, which results in transition midpoints that are in accordance with those found in the global analysis of the transfer efficiency histograms (Supplemental Figure 4). The agreement between these two methods provides additional support for the formation of a ribonucleoprotein complex at low concentrations of RNA.

The RNA concentration dependence of f_{RNP} (Figure 3c, orange) reveals that the midpoint of the transition to the RNP population, $[RNA]_{RNP}$, occurs well below 1 μM RNA for the three longest RNA sequences, whereas for the shortest sequence (RNA1), the RNP population is barely occupied. Furthermore, a comparison across the four titrations indicates that the lengths of the RNA constructs are correlated with their affinity to HCVncd (Figure 3c), but anti-correlated with the extent of compaction they impart on HCVncd in the RNP conformation (Figure 3b). Simply stated, longer RNAs interact more tightly with HCVncd than shorter ones, but give rise to less compaction.

Since the dimensions of the RNP population of HCVncd are so similar to those associated with the FREE population, it is conceivable that this protein remains disordered when associated with the structured RNAs from the viral genome, resulting in a "fuzzy" **[56,57]** ribonucleoprotein complex. To test this hypothesis, we employ the same single-molecule techniques used to characterize the conformational dynamics of the FREE population. For RNA 3, the FREE-to-RNP transition is well separated from the RNP-to-CAP transition, which allows us to work at an RNA concentration that saturates the formation of the RNP population but is insufficient for NLP formation (i.e., 2 μM RNA3). A lifetime density plot (Figure 4a) for this data set reveals that the RNP population also rapidly samples a distribution of FRET efficiencies with a variance ($\sigma^2 = 0.1 \pm 0.02$) well-described by the distance distribution of a WLC, indicating that HCVncd samples a heterogeneous ensemble of conformations even when associated with RNA. Accordingly, nanosecond FCS experiments at the same RNA concentration (Figure 4b) reveal that HCVncd in the RNP population has a reconfiguration time of $\tau_r = 44 \pm 4$ ns, within experimental uncertainty of the $\tau_r = 47 \pm 3$ ns of the FREE population (Figure 2c). In summary, the dimensions of the FREE and RNP populations are remarkably similar, indicating that HCVncd remains unstructured when bound to nucleic acids. This finding support previous suggestions stating that "fuzziness" may be more common in nucleoproteins than previously

thought **[58]**. Furthermore, our nanosecond correlation experiments show that the reconfigurational dynamics of a disordered protein can remain exceedingly rapid and pronounced after RNA binding.

Finally, we utilize fluorescence cross correlation spectroscopy (FCCS) to determine the translational diffusion times (τ_d) of both the FREE and RNP populations of the FRET-labeled HCVncd, which report on the hydrodynamic radius and thus the overall dimensions of these species. An analysis of normalized donor-acceptor cross correlation functions, $g_{DA}(\tau)$ - 1, reveals that τ_d for both populations are nearly identical (Figure 5a; FREE, green; RNP, orange), suggesting that the two have very similar hydrodynamic radii and therefore assemblies significantly larger than the dimer are unlikely to contribute to the RNP population.

2.2.2 Formation of nucleocapsid-like particles at high concentrations of RNA.

At RNA concentrations above ~4 μM, a third population begins to appear in the *E* histograms associated with the three shortest RNA constructs (i.e., RNA1, RNA2, RNA3). This population arises from individual HCVncd subunits within NLPs (i.e., capsomeres, designated as "CAP") that adopt relatively extended conformations. The same analysis as above is used to determine $\langle E \rangle_{CAP}$ and f_{CAP} as a function of RNA concentration (see MATERIALS AND METHODS). The greatly decreased $\langle E \rangle_{CAP} \approx 0.35$ indicates that the $\langle r_{dye\text{-}dye} \rangle$ of HCVncd is about one nanometer larger in the CAP population than it is in the FREE or RNP populations and supports the notion that HCVcp undergoes a large conformational change during NLP assembly **[59]**. Additionally, the RNA-dependence of f_{CAP} (Figure 3c, purple) reveals that the extremely cooperative transition to this expanded population takes place over a very narrow range of concentrations. The increase in f_{CAP} occurs simultaneously with a decrease in f_{RNP} (Figure 3b and 3c), highlighting the existence of an RNA-dependent equilibrium between these two populations.

Again, FCS can be used to gain additional information about the diffusivity of molecules in the CAP conformation. Although the inhomogeneous sample substantially lowers the data quality, it is evident that the HCVncd molecules from the CAP population are associated with large particles that diffuse ~50 times slower (Figure 5a) than either the RNP or FREE species. This observation is consistent with the notion that the CAP population is associated with large, slowly-diffusing NLPs comprised of hundreds of HCVncd capsomeres. Correspondingly, the bursts associated with the CAP conformation have higher photon count rates than those associated with the RNP population (Supplemental Figure

5), reflecting the presence of multiple FRET-labeled particles in each NLP. Note that the 30-fold excess of unlabeled HCVncd included in all experiments is sufficient to prevent intermolecular FRET between the many labeled HCVncd capsomeres within these NLPs. Although it would be desirable to quantitatively characterize the dimensions and dynamics of the CAP conformation using lifetime density plots and nsFCS, high fluorescence anisotropies (> 0.3, Supplemental Figure 6) and the resulting dominance of the rotational contribution to the signal**[60]** severely complicate data analysis and interpretation, specifically regarding quantitative inter-dye distance correlations.

Nevertheless, to gain some kinetic insight into the appearance of the CAP population and the formation of NLPs, we utilize mixing experiments wherein NLP formation is induced by adding RNA to the sample at a final concentration of 20 μM. To accomplish this, we use RNA 1 because at these concentrations the CAP conformation is both well separated from the RNP population and nearly saturated. Prior to the addition of RNA, the short ~1 ms bursts associated with the FREE population of HCVncd give rise to a constant fluorescence signal with relatively small fluctuations when binned at 1 s (Figure 5b, left). After addition of RNA, NLPs begin to form, giving rise to a decrease of the signal associated with the rapidly diffusing FREE and RNP species (Figure 5b, right) as they are incorporated into large slowly diffusing NLPs. This decrease is accompanied by an increase in the signal associated with slowly diffusing NLPs comprised of more extended HCVncd capsomeres. Importantly, the signal from the CAP population is no longer averaged out in the 1 s time bins, giving rise to large fluctuations (i.e., spikes) in the fluorescence signal on this timescale. A single-exponential fit to the decaying signal from the FREE and RNP populations suggests that the NLP assembly process achieves equilibrium with an apparent rate of k(CAP) = 0.59 ± 0.09 min^{-1} (Figure 5b, black line).

Finally, a comparison of the RNA concentrations associated with the transition midpoints (i.e., $[RNA]_{RNP}$ and $[RNA]_{CAP}$) suggests that the length and structure of the RNAs are important factors for efficient NLP assembly. For example, RNA1 (16 nt) is the shortest RNA and forms the least stable complex with HCVncd (i.e., $[RNA]_{RNP}$ = 11 ± 6 μM), however, it is most effective at promoting NLP formation (i.e., $[RNA]_{CAP}$ = 0.9 ± 0.6 μM); because $[RNA]_{CAP} \ll [RNA]_{RNP}$ the RNP population is never highly abundant . Conversely, RNA4 (99 nt) is the longest RNA and forms the most stable complex with HCVncd (i.e., $[RNA]_{RNP}$ = 0.030 ± 0.001 μM), however, it is unable to form NLPs under any of the experimental conditions. It is tempting to propose that the different efficiencies for NLP assembly are

not solely the result of different RNA lengths, but also the result of a context-dependent presentation of the dimer linkage sequence (DLS). This idea is supported by the observation that RNA2 (26 nt), which presents the DLS as a hairpin primed for the formation of a homodimeric kissing complex, is more efficient at inducing NLP formation than RNA3 (27 nt), which incorporates most of the DLS within its stem. Given that NLP formation appears to depends on the context of the DLS, it is conceivable that there is a biochemical link between genome dimerization and nucleocapsid assembly, especially since HCVcp is known to effectively chaperone genome dimerization at the DLS **[14,16]**. Such a relationship between these two processes could explain how the 3`X-RNA is able to act as a molecular switch between biomolecular synthesis and viral assembly **[49-51]**. This would imply that nucleocapsid assembly cannot begin until the host has produced enough HCVcp and genome to effectively chaperone genome dimerization at the otherwise sequestered DLS of the 3`X-RNA. Interestingly, genome dimerization and nucleocapsid assembly are also thought to be intimately related in many retroviral systems **[61,62]**.

2.2.3 Global structure of nucleocapsid-like particles.

To visualize the topological architecture of the large NLPs, single-molecule FRET samples were loaded onto negatively-charged transmission electron microscopy (TEM) grids, stained with uranyl acetate, and imaged at 180,000 × magnification. Given that *Flaviviridae* virus particles are thought to have a triangulation number of T=3 **[63]**, a properly formed nucleocapsid is likely to consist of ~180 HCVncd capsomeres. Because our single-molecule experiments only contain 3 nM HCVncd, one can estimate that the total concentration of NLPs is less than 20 pM. When a few μL of these single-molecule samples are loaded onto TEM grids, it is possible to identify some approximately spherical particles with dimensions that are consistent with the 20-50 nm diameters observed in previous experiments where tRNAs were used to promote NLP formation for a similar HCVcp sequence consisting of the first 82 amino acids **[18]**. However, due to the low surface density (resulting from the low initial HCVncd concentration) it was difficult quantify the dimensions of these NLPs. However, higher concentrations of both HCVncd and RNA (150 nM ulHCVncd; 150 μM RNA1) give rise to TEM grids that are more densely covered with these NLPs (Figure 5C), with an average diameter of $\langle d \rangle$ = 24 ± 4 nm (n = 46). The relatively broad distribution of diameters observed in the TEM images is consistent with previous reports of NLPs lacking a host-derived lipid envelope **[18]** and may be

responsible for heterogeneous translational diffusion observed in our FCCS analysis (Figure 5a, purple). Finally, it is worth noting that the slow diffusivity and low $\langle E \rangle$ of HCVncd molecules from the CAP population is mostly insensitive to the molar ratio of nucleic acid and protein, α = [RNA]/[HCVncd], indicating that the NLPs are not random polyelectrolyte complex coacervates **[64]** but contain a relatively well-defined amount and stoichiometry of the two components.

2.3 A three-state pathway for NLP assembly

The conformational plasticity of HCVncd during RNA-induced nucleocapsid-like particle (NLP) formation can be succinctly summarized by two coupled, RNA dependent, equilibria (Figure 6a). In the absence of RNA, the FREE population of HCVncd is an IDP that rapidly samples a broad ensemble of conformations. At RNA concentrations lower than ~4 μM, the interaction between this positively-charged protein and its viral RNAs give rise to the RNP population, where HCVncd compacts, yet remains disordered and dynamic. Increasing the RNA concentration further promotes a second, highly cooperative, transition to the CAP population. These HCVncd capsomeres appear when hundreds of molecules from the RNP population quickly assemble with free viral RNA hairpins. One important consequence of the intermediate RNP population is that the thermodynamic stability of the CAP population, and thus NLPs, depends on both the concentration and the molar ratio of RNA and HCVncd, α = [RNA]/[HCVncd]. This aspect of the model is most apparent when one increases α by titrating RNA into samples containing 3 nM HCVncd. The abundance of the RNP conformation increases until $\alpha \approx 10^3$, at which point increasing α further begins to favor the CAP population of HCVncd. However, if one then decreases α by greatly increasing the concentration of HCVncd, the model predicts that the additional HCVncd molecules will interact with and sequester the excess free RNA, resulting in conditions where the concentration of free RNA is less than $[RNA]_{CAP}$ and NLP formation is no longer favorable. Indeed, this predicted behavior appears when comparing experiments performed at 20 μM RNA2 with 3 nM ($\alpha \approx 6600$) and 483 nM ($\alpha \approx 40$) HCVncd. In particular, higher concentrations of unlabeled HCVncd shift the equilibrium from the CAP population back towards the more compact, high $\langle E \rangle$, RNP population (Figure 6b). This result highlights that both the absolute concentrations of RNA and protein, and the corresponding molar ratios are important factors for the thermodynamic stability of the NLPs.

3.0 CONCLUSION

Using advanced single-molecule spectroscopic techniques in combination with other biophysical methods, we have shown that the protein domain responsible for forming the HCV nucleocapsid (HCVncd) is a typical IDP that rapidly samples a broad ensemble of conformations. HCVncd can form a complex with small RNA structures from the HCV genome at concentrations below a few μM, resulting in a slightly more compact, yet still very dynamic ensemble of disordered conformations. At higher RNA concentrations (e.g., 20 μM), these ribonucleoprotein complexes assemble with additional RNA on the timescale of minutes, giving rise to a distribution of large, slowly diffusing, nucleocapsid-like particles (NLPs), each of which contains many individual HCVncd capsomeres. Within the NLPs, the HCVncd capsomeres adopt relatively expanded conformations. These results highlight the conformational plasticity of HCVncd during the assembly of NLPs and yield a model for NLP assembly that consists of two coupled, RNA dependent, equilibria. Finally, various genomic RNA constructs from the 3′ X-RNA of HCV differentially induce NLP formation, depending on the structural context of the dimer linkage sequence (DLS). If genome dimerization is indeed a prerequisite for nucleocapsid formation, then context dependent DLS-mediated dimerization may explain how the 3′ X-RNA could act as molecular switch to turn off viral replication and induce particle assembly.

4.0 MATERIAL AND METHODS

4.1 Expression, labeling, and purification of HCVncd

The PFAM **[65]** database was used to identify a suitable variant (GenBank: CAE46584.1 **[21]**) of the nucleocapsid domain of the hepatitis C virus core protein (HCVncd) that contains a minimal number of cysteine residues (which interfere with labelling) and aromatic residues (which can quench single-molecule fluorophores). Four mutations were introduced into this sequence in order to produce a single-molecule FRET-viable HCVncd protein: S2C, T65C, C98M, and ΔW134 (Supplemental Figure 1). The corresponding gene was codon optimized for *E.coli*, cloned into a pET-47b(+) vector, and transfected into BL21 (DE3) cells. Recombinant expression was performed in accordance with existing protocols **[13,22,42]** with slight modifications. Briefly, cells were grown in 2 L of LB media for 3 h at 37 °C. Expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), followed by an additional 3 h growth period at 37 °C. Cells were pelleted, lysed, and homogenized in denaturing buffer containing 6 M guanidinium chloride (GdmCl). After the insoluble material was removed from the homogenate, His-tagged HCVncd was purified using denaturing immobilized metal

ion affinity chromatography (IMAC). The His-tag was enzymatically cleaved using recombinantly expressed 3C protease, at which point HCVncd was isolated from the His-tag and the protease via a second round of denaturing (6 M GdmCl) IMAC. Unlabeled HCVncd was further purified using a high performance liquid chromatography (HPLC) instrument (1100 Series, Agilent; Santa Clara, CA) and a reversed-phase column (Reprosil Gold 200 C18 5 μm, Dr. Maisch HPLC GmbH; Ammerbuch-Entringen, Germany) using a water/acetonitrile gradient containing 0.1% trifluoroacetic acid. The protein mass, and thus peptide sequence, was confirmed by electrospray ionization mass spectrometry (Functional Genomics Center, University of Zurich; Zurich, Switzerland). Purified HCVncd was buffer exchanged into 50 mM sodium phosphate pH 7.0 with 40 mM KCl and stored at -20 °C. After reduction with tris(2-carboxyethyl)phosphine (TCEP), HCVncd was allowed to reacted with 5× molar equivalents of both Alexa 488-maleimide (donor) and Alexa 594-maleimide (acceptor) under denaturing conditions (6 M GdmCl). The HPLC elution profiles confirmed that all of the protein was coupled to fluorophores. However, the similar retention times for the donor-donor, donor-acceptor, acceptor-donor, and acceptor-acceptor species prevented complete separation of the FRET-labeled HCVncd populations (Supplemental Figure 2). However, residual populations of the differently labeled species can be useful for determining correction factors such as direct excitation, cross-talk, relative detection efficiencies, and other photophysical parameters like the intrinsic lifetime of the donor (τ_D) and acceptor (τ_A) fluorophores **[66]**. The FRET-labeled HCVncd was lyophilized and re-suspended to a final concentration of 10 nM in storage buffer (50 mM sodium phosphate buffer pH 7.0, 10% v/v glycerol, 1 M GdmCl, and 0.1% Tween-20) before being split into 50 μl aliquots and stored at -80 °C until use. We note that our non-native IMAC and HPLC purification procedures do not affect HCVncp activity, as demonstrated by its ability to bind RNA **[14-16]** and form nucleocapsid-like particles **[13,18]**, and as expected for IDPs in view of their flat and smooth free energy surfaces **[67]**.

4.2 Circular Dichroism Spectroscopy

Unlabeled HCVncd was brought to a final concentration of 1.4 μM in a 50 mM sodium phosphate buffer (pH 7.0) containing 40 mM KCl and measured at room temperature on a spectropolarimeter (J-810, Jasco; Oklahoma City, OK) in a cylindrical quartz cell with a 0.5 mm path length. Data were acquired by scanning the circularly polarized light from 185 nm to 250 nm over the course of 10

minutes; this process was repeated for a total of 8 hours. The resulting spectrum was analyzed using the DichroWeb online analysis tool **[23]**.

4.3 Single-Molecule Fluorescence Instrument

Single-molecule fluorescence experiments were performed on modified version of a commercially-available, inverted confocal microscope system (MicroTime 200, PicoQuant; Berlin, Germany) operated in an epifluorescence configuration. Briefly, a 1.2 NA 60× water immersion objective (UPLSAPO 60XW, Olympus; Tokyo, Japan) was used to focus the light from the excitation sources to a diffraction-limited spot. For direct excitation of the donor, a diode laser (LDH-D-C-485, PicoQuant; Berlin, Germany) was operated at either 488 nm in continuous wave mode or at 485 nm in pulsed mode (20 MHz with a pulse width < 550 ps). For direct excitation of the acceptor, the broadband output of a pulsed (20 MHz with a pulse width < 10 ps) super continuum laser (SC450-PP, Fianium; Southampton, United Kingdom) was filtered to ~ 585 nm using a band pass filter (z582/15, Chroma; Bellows Falls, VT). The same objective collects the emitted fluorescence, which is directed through a 100 μm confocal pinhole via a dual band dichroic beam splitter (z485/568, Chroma; Bellows Falls, VT), before being separated first by polarization and then by color (585 DCXR, Chroma; Bellows Falls, VT), resulting in a total of four photon detection streams (i.e., horizontally and vertically polarized light for both donor and acceptor channels). Additionally, band pass filters were used to reject unwanted fluorescence from both the donor (ET525/50M, Chroma; Bellows Falls, VT) and accepter (HQ650/100, Chroma; Bellows Falls, VT) channels. The arrival times of individual fluorescence photons detected by single-photon avalanche diodes (τ-SPAD fast, PicoQuant; Berlin, Germany) were recorded by a time-correlated single-photon counting module (HydraHarp 400, PicoQuant; Berlin, Germany) with 16 ps resolution prior to being saved for analysis using an in-house software package and Mathematica (version 10.4, Wolfram Research). All single-molecule measurements were conducted in phosphate-buffered saline containing 50 mM sodium phosphate buffer at pH 7.0, 80 mM NaCl (unless noted otherwise), 5 mM GdmCl (unless noted otherwise), 0.01% w/v Tween 20, 143 mM β-mercaptoethanol, and 3 nM unlabeled HCVncd. Despite being a mild detergent, Tween 20 has no detectable effect on the conformational dimensions or dynamics of HCVncd. It only serves to inhibit the absorption of proteins onto the surfaces of the sample cell, thereby enabling prolonged measurements. The focus was set 50 μm into solution above a

Polyethylene glycol-coated quartz coverslip (PEG_02, Microsurfaces, Inc.; Englewood, NJ), which was used to minimize protein adsorption.

4.4 Single-Molecule Fluorescence Time Traces

Because the arrival time of each and every detected photon is recorded with 16-ps precision, the four photon steams can be binned with any desired choice of bin size, resulting in a fluorescence time trace that can be easily visualized. A convenient bin size for free diffusion experiments is 1 ms, which is large enough to encompass a single molecule diffusing through the detection volume without including either additional background fluorescence or neighboring bursts. A larger bin time of 1 s can be used to visualize slower dynamics (Figure 5b), like the translational diffusion of nucleocapsid-like particles (NLPs) or a change in the solution concentration of rapidly diffusing species (e.g., free and RNA-bound HCVncd).

4.5 Single-Molecule Transfer Efficiency Histograms

When using pulsed interleaved excitation **[41]** (PIE; 485 nm and 582/15 nm), individual bursts were identified by binning the record of photons at 1 ms and identifying bins with more than 75 photons (after correcting for background, differential excitation and detection efficiencies of the two dyes, cross talk, and direct excitation) resulting from either donor or acceptor direct excitation. Only bursts with a fluorescence stoichiometry **[66]** (S = signal after donor excitation/signal after both donor and acceptor excitation) between 0.3 and 0.7 were considered for analysis to prevent complications associated with the donor-donor and acceptor-acceptor species that were not fully removed during HPLC purification (Supplemental Figure 2). Multiple experimental parameters like fluorescence lifetime, fluorescence anisotropy, and ratiometric transfer efficiency (E) are calculated for each 1-ms burst, which are displayed as a histogram to show the frequency of bursts detected over a specific range of transfer efficiencies, for example. The mean transfer efficiency, $\langle E \rangle$, of the single distributions were determined by fitting the histograms to a log-normal distribution. To convert the experimentally determined $\langle E \rangle$ into a reliable mean inter-dye distance, $\langle r_{dye\text{-}dye} \rangle$, one must ensure that: (1) the relative orientation of the two fluorophores is isotropic and, (2) the distribution of inter-dye distance, $r_{dye\text{-}dye}$, is sampled on a timescale that is between the mean interphoton time (~ 10 μs) and the fluorescence lifetime of the donor (~ 4 ns), both of which are valid in this case. As a result, $\langle E \rangle = \int_0^\infty dr\, E(r_{dye\text{-}dye})\, P(r_{dye\text{-}dye})$ with

$E(r_{dye\text{-}dye}) = 1/(1+(r_{dye\text{-}dye}/R_0)^6)$, where R_0 = 5.4 nm is the Förster Radius for this donor-acceptor FRET pair. For this work we choose to use a worm-like chain **[68]** to describe the distribution of inter-dye distances,

$$P(r_{dye\text{-}dye}) = \frac{4\pi N r_{dye\text{-}dye}^2}{l_c^2\left[1-\left(r_{dye\text{-}dye}/l_c\right)^2\right]^{9/2}} exp\left(-\frac{3l_c}{4l_p\left[1-\left(r_{dye\text{-}dye}/l_c\right)^2\right]}\right),$$

where l_p and l_c are the persistence and contour length of the chain, and N is a normalization constant. For HCVncd, we assume l_c = 274 Å, which corresponds to 3.8 Å per residue**[69]** plus an additional 1.54 Å for each of the 10 bonds in the linkers connecting the dyes to the cysteine.

Following the work of Hofmann *et al*. **[28]** and Aznauryan *et al* **[42]**, we use the relationship $\langle r_{dye\text{-}dye}^2\rangle^{1/2} = r_0 N^{\nu}$ to calculate the scaling exponent for HCVncd, where the number of segments, N, contains 63 amino acid segments plus an additional 9 segments to account for the linkers that attach the fluorophores to the protein. The prefactor, $r_0 = \sqrt{(2l_p b)} \approx 6.0$ angstroms, is determined using our experimentally determined persistence length, l_p = 4.8 angstroms, and a peptide segment length of b = 3.8 angstroms, corresponding to the well-documented Cα- Cα distance of extended polypeptides This relationship give rise to a scaling exponent of $\nu \approx 0.49$, which is very near the value of ½ expected for a polymer under theta conditions.

To determine $\langle E\rangle$ and the fractional abundance (f) of each population as a function of a given experimental variable, E histograms were globally fit to a sum of three log-normal distribution corresponding to the FREE, RNP, and CAP conformations, where peak positions, widths, and asymmetries were shared parameters. The two coupled, RNA-dependent, equilibria (Figure 6a) are each assumed to be well described by simple cooperative binding models:

$$K_{\text{FREE-RNP}} = \frac{[\text{RNA}]^{n1}}{[RNA]_{RNP}^{n1}} \text{ and } K_{\text{RNP-CAP}} = \frac{[\text{RNA}]^{n2}}{[RNA]_{CAP}^{n2}}, \quad (1)$$

where $[RNA]_{RNP}$ and $[RNA]_{CAP}$ are the RNA concentrations required to establish an equilibrium constant of one and $n1$ and $n2$ measure how strongly the equilibrium constant depends on the RNA concentration (i.e., $\frac{\partial \ln K_{\text{FREE-RNP}}}{\partial \ln[RNA]}$ and $\frac{\partial \ln K_{\text{RNP-CAP}}}{\partial \ln [RNA]}$) **[70]**. The fractional abundance, f, of both states can be described in terms of these two equilibrium constants as a function of RNA concentration (Figure 3c):

$$f_{RNP} = \frac{K_{FREE\text{-}RNP}}{1 + K_{FREE\text{-}RNP} + (K_{FREE\text{-}RNP}\, K_{RNP\text{-}CAP})} \text{ and } f_{CAP} = \frac{K_{FREE\text{-}RNP}\, K_{RNP\text{-}CAP}}{1 + K_{FREE\text{-}RNP} + (K_{FREE\text{-}RNP}\, K_{RNP\text{-}CAP})}. \quad (2)$$

Because of the shared parameters, both f_{RNP} and f_{CAP} are fit simultaneously.

4.6 Lifetime vs E Density Plots (Photon Delay Time vs *E*)

A burst identification procedure, similar to the one describe above, was used to identify bursts, except that a more stringent threshold of 100 photons per burst was used in order to better determine the mean excitation-emission delay time of donor, $\langle t_D \rangle = \tau_D$, and acceptor, $\langle t_A \rangle = \tau_A$, photons within each burst. The donor lifetime in the absence of the acceptor (τ_D^0) was determined by fitting the fluorescence decay histogram of the donor-only species ($S > 0.9$) to a convolution of the instrument response function and a single exponential decay. These values differed by less than 5 % from those obtained by fitting the tail of this histogram to a single exponential decay. The acceptor lifetime in the absence of the donor (τ_A^0), was obtained using the same procedure to fit the acceptor-only species ($S < 0.15$) after direct excitation. The mean excitation-emission delay time for donor photons can be used to calculate the relative mean donor lifetime (i.e., τ_D/τ_D^0) of the FRET-labeled population (i.e., $0.3 < S < 0.7$). As previously shown**[38,39]**, a similar quantity can be calculated using the mean excitation-emission delay time of acceptor photons (i.e., $(\tau_A - \tau_A^0)/\tau_D^0 = \tau_D^A/\tau_D^0$). For systems with a single fixed FRET efficiency, these two quantities are equal and related to *E* via:

$$\tau_D/\tau_D^0 = \tau_D^A/\tau_D^0 = 1\text{-}E, \quad (3)$$

which corresponds to the diagonal black line in Figure 2c. When transfer efficiency fluctuations occur on a timescale between the donor fluorescence lifetime (~4 ns) and the burst duration (~1 ms), events will cluster above (τ_D/τ_D^0 vs *E*) or below (τ_D^A/τ_D^0 vs *E*) this line (Figure 2c, blue and orange, respectively). Quantitatively, the vertical displacement from the diagonal line is related to the variance of the underlying FRET efficiency distribution, σ^2, via

$$\tau_D/\tau_D^0 = (1\text{-}E) + \sigma^2/(1\text{-}E) \quad \text{and} \quad \tau_D^A/\tau_D^0 = (1\text{-}E) - \sigma^2/E. \quad (4)$$

The curved lines above and below the diagonal of the lifetime density plots correspond to the expected behavior for a rapidly sampled $r_{dye\text{-}dye}$ distribution determined by a worm-like chain (WLC) with a fixed contour length and a variable persistence length (l_p) used to modulate the mean inter-dye distance, $\langle r_{dye\text{-}dye} \rangle$, as a function of experimental conditions.

4.7 Single-Molecule FRET Fluorescence Cross Correlation Spectroscopy.

The donor-acceptor cross correlation functions, $g(\tau)$, were generated using photon lag times (τ) between 10^{-6} s and 10^{1} s and analyzed using a model for three-dimensional diffusion through a Gaussian observation volume in order to determine the translational diffusion time **[71]**. For nanosecond FCS experiments, the donor excitation source was operated in continuous wave mode. Only photon lag times between 0 and 150 ns are included in the correlation analysis. The resulting correlation functions, $g(\tau)$, for the donor-donor, donor-acceptor, and acceptor-acceptor signal were globally fit with two common time constants: a negative-amplitude (anti-correlated) low-ns (< 10 ns) component (corresponding to the antibunching behavior of single emitters), and a slower (> 10 ns) component for protein chain dynamics that has a negative amplitude (anti-correlated) only for the donor-acceptor correlation.

4.8 Transmission Electron Microscopy

All transmission electron microscopy images were acquired at the Center for Microscopy and Image Analysis (University of Zurich; Zurich, Switzerland). Briefly, negatively charged TEM grids were inverted and placed, for a duration of three minutes, onto a 20 μL drop of sample where experimental conditions favor NLP formation. Low particle density conditions were: 100 pM labeled HCVncd, 3 nM unlabeled HCVncd, 20 μM RNA1, whereas for high particle density conditions were: 100 pM labeled HCVncd, 150 nM unlabeled HCVncd, 150 μM RNA1. Then, the inverted TEM grids were placed on top of 20 μL drops of water for three minutes to wash away any loosely bound material. Finally, the inverted TEM grids were placed on top of 20 μL drop of 2% uranyl acetate for three minutes to enhance image contrast. TEM grids were imaged on a 100 kV microscope (TEM CM100, Phillips; Amsterdam, The Netherlands), and the associated digital images were acquired using a charge-coupled device (Orius 1000, Gatan; Pleasanton, CA).

5.0 ACCESSION NUMBERS

The amino acid sequence for the hepatitis C virus core protein can be found in GenBank under accession number CAE46584.

6.0 ACKNOWLEDGEMENT

Support for this work has been provided by the Swiss National Science Foundation (B.S.) and the European Molecular Biology Organization (grant number ATLF 471-2015 to E.D.H). Funding for open access charge: Swiss National Science Foundation. We would also like to thank the Center for Microscopy and Image analysis at the University of Zurich for their assistance and support with Transmission Electron Microscopy.

7.0 CONFLICT OF INTEREST

None Declared.

8.0 FIGURES AND FIGURE CAPTIONS

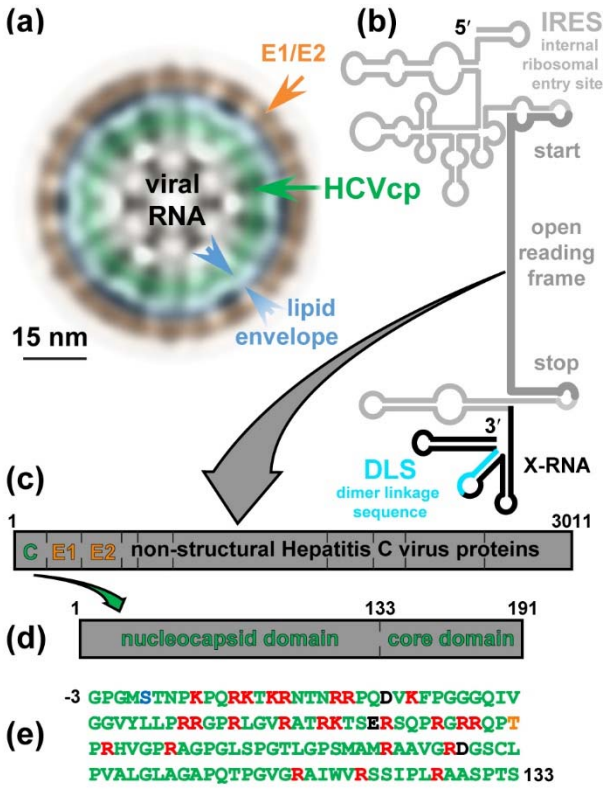


Figure 1: Architecture of the hepatitis C virus. (a) Cross section of a cryo-electron microscopy-based 3D reconstruction of a HCV-like particle (figure adapted from ref. **[3]**) colored to show the multilayer structural organization, which includes: (*i*) the outer shell (orange) made up of envelope proteins (E1/E2), (*ii*) the lipid envelope (blue), and (*iii*) the nucleocapsid (green). Of particular interest to this study is the hepatitis C virus core protein (HCVcp, green), which interacts with the genomic RNA to form a nucleocapsid. (b) Schematic representation of the positive-sense single-stranded RNA genome depicting relevant sequence elements associated with viral replication and assembly. Domain maps of (c) the viral polyprotein and (d) the HCVcp. (e) Amino acid sequence of core protein's nucleocapsid domain (HCVncd), highlighting the positive (red) and negative (black) charges along the polypeptide. The blue and orange amino acids represent the location of the cysteine residues introduced for coupling of donor (Alexa 488) and acceptor (Alexa 594) fluorophores.

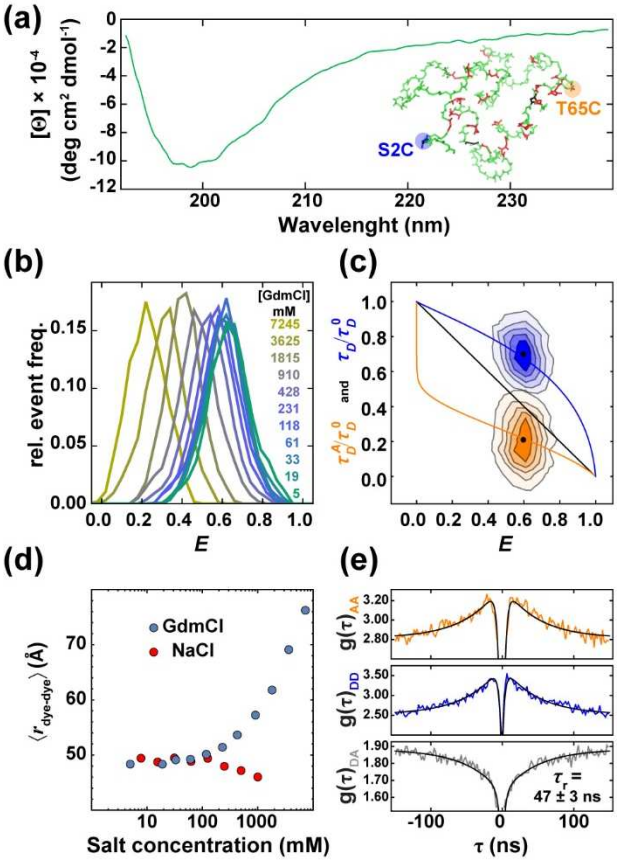


Figure 2: Intrinsic disorder within HCVncd. (a) Circular dichroism spectrum of HCVncd in non-denaturing buffer (green line). The inset shows a backbone representation of HCVncd colored according to Figure 1e. (b) Transfer efficiency (*E*) histograms for single HCVncd molecules in buffered solutions with increasing (green → blue → orange) concentrations of guanidinium chloride (GdmCl); higher concentrations shift ⟨*E*⟩ to lower values. (c) Two-dimensional lifetime density plots of τ_D/τ_D^0 (blue) or τ_D^A/τ_D^0 (orange) photons (see MATERIALS AND METHODS for details). The solid black line indicates the expected dependence for a single fixed FRET efficiency; blue (τ_D/τ_D^0) and orange (τ_D^A/τ_D^0) lines correspond to the expected dependence for a worm-like chain (WLC) that rapidly samples its distance distribution; black points correspond to the average values obtained by combining all bursts. (d) Plot of $\langle r_{\text{dye-dye}} \rangle$ vs salt concentration for both GdmCl (blue) and NaCl (red) highlighting the corresponding changes in dimensions. The addition of NaCl minimizes electrostatic repulsion, allowing HCVncd to sample more compact conformations. The addition of GdmCl leads to an expansion of the peptide chain that outcompetes the compaction driven by charge screening. (e) Donor-donor ($g(\tau)_{DD}$, blue), donor-acceptor ($g(\tau)_{DA}$, grey) and acceptor-acceptor ($g(\tau)_{AA}$, orange) nanosecond cross correlation functions from diffusing FRET-labeled HCVncd molecules in non-denaturing buffer. A global fit (black lines) to all three correlation functions using a common decay time shows that a reconfiguration time of τ_r = 47 ± 3 ns can describe the correlated components of $g(\tau)_{DD}$ and $g(\tau)_{AA}$, as well as the corresponding anti-correlated component of $g(\tau)_{DA}$. The sharp decrease in $g(\tau)$ at inter-photon times (τ) approaching 0 ns results from the characteristic anti-bunching behavior of single emitters.

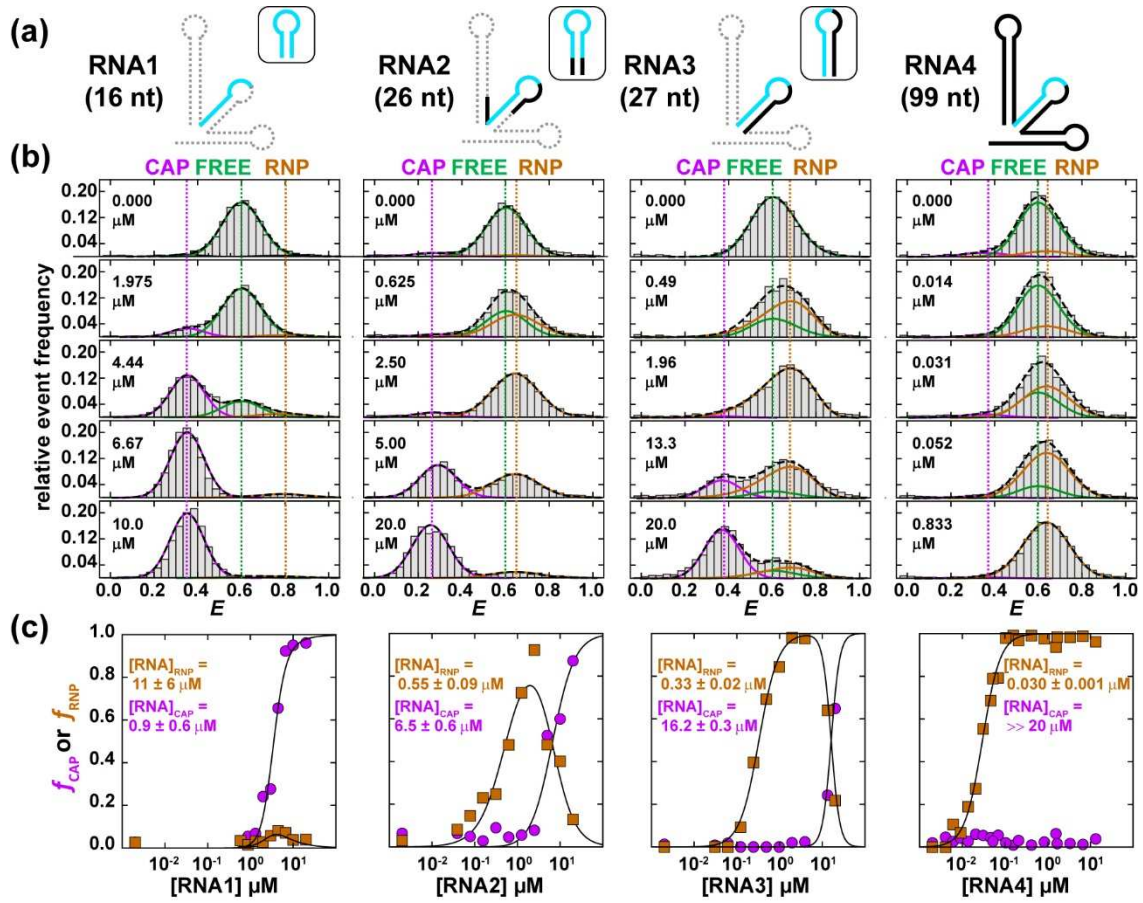


Figure 3: Interactions between HCVncd and segments of the 3′ X-RNA. (a) Secondary structure diagrams of the four RNA sequences from the viral genome (Figure 1b) used to induce nucleocapsid formation. Dashed lines correspond to regions of the 3′ X-RNA that were excised in each of the four nucleic acids; the cyan line represents the dimer linkage sequence (DLS). Insets show the predicted secondary structure **[72]** for each of the three truncated sequences. (b) Example *E* histograms of HCVncd at RNA concentrations ranging from 0 to 20 μM. At concentrations below ~4 μM, the FREE population (green) binds to RNA, resulting in the more compact RNP population (orange). At higher RNA concentrations, this ribonucleoprotein intermediate begins to form nucleocapsid-like particles, wherein the individual HCVncd capsomeres sample an expanded, low ⟨*E*⟩, CAP population (purple). All *E* histograms within a single RNA titration series are globally fit to a sum of three log-normal functions, i.e. using shared peak parameters. (c) Fits are used to determine the fractional abundance, *f*, of each population as a function of RNA concentration. Solid lines represent fits to cooperative binding functions (see MATERIALS AND METHODS).

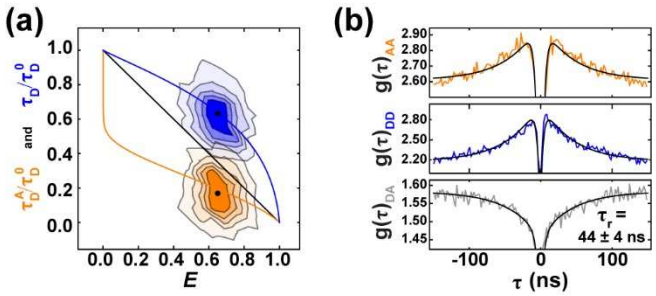


Figure 4: Conformational dynamics of the RNP population. (a) Lifetime density plot for FRET-labeled HCVncd molecules in the presence of 2 μM RNA3 where the RNP population is most abundant. Colors and lines are the same as in Figure 2c. (b) Nanosecond FCS cross correlation functions. The donor-donor ($g(\tau)_{DD}$, blue), donor-acceptor ($g(\tau)_{DA}$, grey) and acceptor-acceptor ($g(\tau)_{AA}$, orange) correlations are globally fit (black lines) to measure the correlated component of $g(\tau)_{DD}$ and $g(\tau)_{AA}$, as well as the anti-correlated component of $g(\tau)_{DA}$. This analysis reveals a reconfiguration time of τ_r = 44 ± 4 ns associated with $r_{dye\text{-}dye}$ fluctuations.

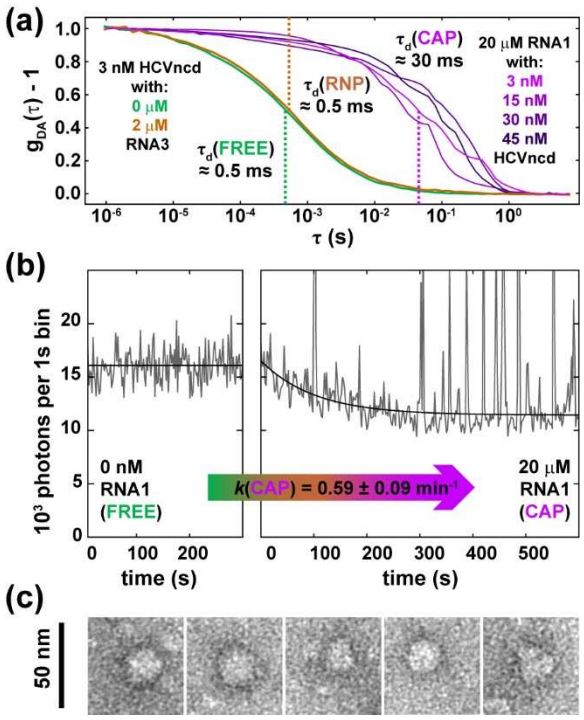


Figure 5: RNA-induced nucleocapsid-like particle formation of HCVncd. (a) Normalized donor-acceptor cross correlation functions, $g_{DA}(\tau)$ - 1, showing the translational diffusion of FRET-labeled HCVncd at conditions where the FREE (green), RNP (orange), and CAP (purple) populations dominate. (b) Confocal fluorescence recordings of diffusing HCVncd molecules before (left) and after (right) addition of 20 μM RNA1. Prior to the addition of RNA, the ~1 ms bursts associated with the FREE population of HCVncd are averaged out in each of the 1 s time bins. After addition of RNA, NLPs begin to form, resulting in a decrease in the signal associated with the FREE and RNP species and a concurrent increase in the abundance of large spikes of fluoresce associated with the CAP population of HCVncd. Importantly, these spikes have much higher photon count rates than the bursts associated with the FREE and RNP populations and are no longer averaged out in the 1 s time bins. The accumulation of the CAP population takes place with an apparent rate constant of k(CAP) = 0.59 ± 0.09 min^{-1}. (c) Transmission electron microscopy images of NLPs formed using FRET-labeled and unlabeled HCVncd and RNA1. Images were taken at 180,000 × magnification; scale bar is 50 nm.

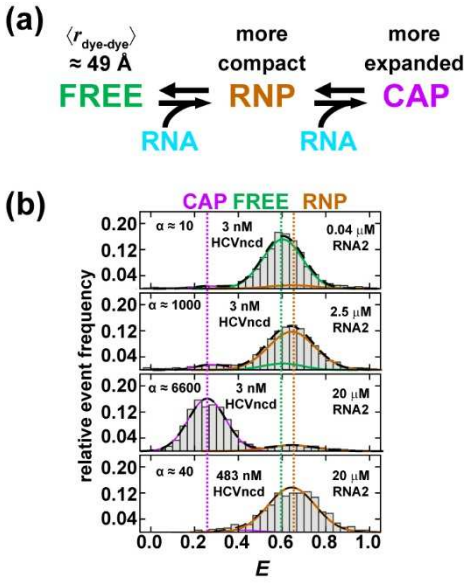


Figure 6: Model for nucleocapsid-like particle assembly. (a) The FREE population of HCVncd is comprised of IDPs that rapidly sample a broad distribution of FRET efficiencies, which can be well described by an unstructured chain. This conformational ensemble interacts with genomic RNAs at concentrations below a few micromolar, resulting in the more compact yet still dynamic and disordered, RNP population of HCVncd. Higher RNA concentrations promote a second, much more abrupt, conformational transition wherein the ribonucleoprotein complexes from the RNP population assemble with free RNA to form large nucleocapsid-like particles (NLPs). These NLPs contain many individual HCVncd capsomeres, which have a much larger $\langle r_{\text{dye-dye}} \rangle$. (b) *E* histograms at various molar ratios of RNA to protein (i.e., α = [RNA]/[HCVncd]). Increasing α first populates the RNP population and then the CAP population. However, increasing the concentration of unlabeled HCVncd, thus decreasing α, shifts the equilibrium back towards the higher $\langle E \rangle$ population associated with the RNP population.

9.0 REFERENCES

1. Yamane, D., McGivern, D.R., Masaki, T. & Lemon, S.M. (2013) Liver injury and disease pathogenesis in chronic hepatitis C. Curr Top Microbiol, 369, 263-288.

2. Perz, J.F., Armstrong, G.L., Farrington, L.A., Hutin, Y.J. & Bell, B.P. (2006) The contributions of hepatitis B virus and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. J Hepatol, 45, 529-538.

3. Yu, X., Qiao, M., Atanasov, I., Hu, Z., Kato, T., Liang, T.J. & Zhou, Z.H. (2007) Cryo-electron microscopy and three-dimensional reconstructions of hepatitis C virus particles. Virology, 367, 126-134.

4. Kato, N. (2000) Genome of human hepatitis C virus (HCV): gene organization, sequence diversity, and variation. Microb Comp Genomics, 5, 129-151.

5. Scheel, T.K. & Rice, C.M. (2013) Understanding the hepatitis C virus life cycle paves the way for highly effective therapies. Nat Med, 19, 837-849.

6. Penin, F., Dubuisson, J., Rey, F.A., Moradpour, D. & Pawlotsky, J.M. (2004) Structural biology of hepatitis C virus. Hepatology, 39, 5-19.

7. Gawlik, K. & Gallay, P.A. (2014) HCV core protein and virus assembly: what we know without structures. Immunol Res, 60, 1-10.

8. Duvignaud, J.B., Savard, C., Fromentin, R., Majeau, N., Leclerc, D. & Gagne, S.M. (2009) Structure and dynamics of the N-terminal half of hepatitis C virus core protein: an intrinsically unstructured protein. Biochem Biophys Res Commun, 378, 27-31.

9. Fan, X., Xue, B., Dolan, P.T., LaCount, D.J., Kurgan, L. & Uversky, V.N. (2014) The intrinsic disorder status of the human hepatitis C virus proteome. Mol Biosyst, 10, 1345-1363.

10. Xue, B., Blocquel, D., Habchi, J., Uversky, A.V., Kurgan, L., Uversky, V.N. & Longhi, S. (2014) Structural disorder in viral proteins. Chem Rev, 114, 6880-6911.

11. Brucale, M., Schuler, B. & Samori, B. (2014) Single-molecule studies of intrinsically disordered proteins. Chem Rev, 114, 3281-3317.

12. Ferreon, A.C., Moran, C.R., Gambin, Y. & Deniz, A.A. (2010) Single-molecule fluorescence studies of intrinsically disordered proteins. Methods Enzymol, 472, 179-204.

13. Kunkel, M., Lorinczi, M., Rijnbrand, R., Lemon, S.M. & Watowich, S.J. (2001) Self-assembly of nucleocapsid-like particles from recombinant hepatitis C virus core protein. J Virol, 75, 2119-2129.

14. Ivanyi-Nagy, R., Kanevsky, I., Gabus, C., Lavergne, J.P., Ficheux, D., Penin, F., Fosse, P. & Darlix, J.L. (2006) Analysis of hepatitis C virus RNA dimerization and core-RNA interactions. Nucleic Acids Res, 34, 2618-2633.

15. Ivanyi-Nagy, R., Lavergne, J.P., Gabus, C., Ficheux, D. & Darlix, J.L. (2008) RNA chaperoning and intrinsic disorder in the core proteins of Flaviviridae. Nucleic Acids Res, 36, 712-725.

16. Cristofari, G., Ivanyi-Nagy, R., Gabus, C., Boulant, S., Lavergne, J.P., Penin, F. & Darlix, J.L. (2004) The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral(+) strand RNA in vitro. Nucleic Acids Res, 32, 2623-2631.

17. Dolan, P.T., Roth, A.P., Xue, B., Sun, R., Dunker, A.K., Uversky, V.N. & LaCount, D.J. (2015) Intrinsic disorder mediates hepatitis C virus core-host cell protein interactions. Protein Sci, 24, 221-235.

18. Majeau, N., Gagne, V., Boivin, A., Bolduc, M., Majeau, J.A., Ouellet, D. & Leclerc, D. (2004) The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation. J Gen Virol, 85, 971-981.

19. Boulant, S., Montserret, R., Hope, R.G., Ratinier, M., Targett-Adams, P., Lavergne, J.P., Penin, F. & McLauchlan, J. (2006) Structural determinants that target the hepatitis C virus core protein to lipid droplets. J Biol Chem, 281, 22236-22247.

20. Lindenbach, B.D. & Rice, C.M. (2013) The ins and outs of hepatitis C virus entry and assembly. Nat Rev Microbiol, 11, 688-700.

21. Colina, R., Casane, D., Vasquez, S., Garcia-Aguirre, L., Chunga, A., Romero, H., Khan, B. & Cristina, J. (2004) Evidence of intratypic recombination in natural populations of hepatitis C virus. J Gen Virol, 85, 31-37.

22. Boulant, S., Vanbelle, C., Ebel, C., Penin, F. & Lavergne, J.P. (2005) Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. J Virol, 79, 11353-11365.

23. Whitmore, L. & Wallace, B.A. (2008) Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. Biopolymers, 89, 392-400.

24. Janes, R.W. (2008) In Wallace, B. A. and Janes, R. W. (eds.). IOS Press.

25. Abdul-Gader, A., Miles, A.J. & Wallace, B.A. (2011) A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy. Bioinformatics, 27, 1630-1636.

26. Provencher, S.W. & Glockner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. Biochemistry, 20, 33-37.

27. Muller-Spath, S., Soranno, A., Hirschfeld, V., Hofmann, H., Ruegger, S., Reymond, L., Nettels, D. & Schuler, B. (2010) Charge interactions can dominate the dimensions of intrinsically disordered proteins. P Natl Acad Sci USA, 107, 14609-14614.

28. Hofmann, H., Soranno, A., Borgia, A., Gast, K., Nettels, D. & Schuler, B. (2012) Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. P Natl Acad Sci USA, 109, 16155-16160.

29. Soranno, A., Buchli, B., Nettels, D., Cheng, R.R., Muller-Spath, S., Pfeil, S.H., Hoffmann, A., Lipman, E.A., Makarov, D.E. & Schuler, B. (2012) Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. P Natl Acad Sci USA, 109, 17800-17806.

30. Mao, A.H., Crick, S.L., Vitalis, A., Chicoine, C.L. & Pappu, R.V. (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. P Natl Acad Sci USA, 107, 8183-8188.

31. Borgia, A., Zheng, W., Buholzer, K., Borgia, M.B., Schuler, A., Hofmann, H., Soranno, A., Nettels, D., Gast, K., Grishaev, A. *et al.* (2016) Consistent View of Polypeptide Chain Expansion in Chemical Denaturants from Multiple Experimental Methods. J Am Chem Soc, 138, 11714-11726.

32. Haran, G. (2012) How, when and why proteins collapse: the relation to folding. Curr Opin Struct Biol, 22, 14-20.

33. Schuler, B. & Eaton, W.A. (2008) Protein folding studied by single-molecule FRET. Curr Opin Struct Biol, 18, 16-26.

34. Schuler, B., Soranno, A., Hofmann, H. & Nettels, D. (2016) Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins. Annu Rev Biophys, 45, 207-231.

35. Gopich, I.V. & Szabo, A. (2007) Single-molecule FRET with diffusion and conformational dynamics. J Phys Chem B, 111, 12925-12932.

36. Kalinin, S., Valeri, A., Antonik, M., Felekyan, S. & Seidel, C.A. (2010) Detection of structural dynamics by FRET: a photon distribution and fluorescence lifetime analysis of systems with multiple states. J Phys Chem B, 114, 7983-7995.

37. Laurence, T.A., Kong, X., Jager, M. & Weiss, S. (2005) Probing structural heterogeneities and fluctuations of nucleic acids and denatured proteins. P Natl Acad Sci USA, 102, 17348-17353.

38. Gopich, I.V. & Szabo, A. (2012) Theory of the energy transfer efficiency and fluorescence lifetime distribution in single-molecule FRET. P Natl Acad Sci USA, 109, 7747-7752.

39. Chung, H.S., Louis, J.M. & Gopich, I.V. (2016) Analysis of Fluorescence Lifetime and Energy Transfer Efficiency in Single-Molecule Photon Trajectories of Fast-Folding Proteins. J Phys Chem B, 120, 680-699.

40. O'Connor, D. (2012) *Time-correlated single photon counting*. Academic Press.

41. Hendrix, J. & Lamb, D.C. (2013) Pulsed Interleaved Excitation: Principles and Applications. Methods Enzymol, 518, 205-243.

42. Aznauryan, M., Delgado, L., Soranno, A., Nettels, D., Huang, J.R., Labhardt, A.M., Grzesiek, S. & Schuler, B. (2016) Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. Proc Natl Acad Sci U S A, 113, E5389-5398.

43. Hoffmann, A., Kane, A., Nettels, D., Hertzog, D.E., Baumgartel, P., Lengefeld, J., Reichardt, G., Horsley, D.A., Seckler, R., Bakajin, O. *et al.* (2007) Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. 104, 105-110.

44. Soranno, A., Holla, A., Dingfelder, F., Nettels, D., Makarov, D.E. & Schuler, B. (2017) Integrated view of internal friction in unfolded proteins from single-molecule FRET, contact quenching, theory, and simulations. P Natl Acad Sci USA, 114, E1833-E1839.

45. Zheng, W., Borgia, A., Buholzer, K., Grishaev, A., Schuler, B. & Best, R.B. (2016) Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment. 138, 11702-11713.

46. Nettels, D., Hoffmann, A. & Schuler, B. (2008) Unfolded protein and peptide dynamics investigated with single-molecule FRET and correlation spectroscopy from picoseconds to seconds. J Phys Chem B, 112, 6137-6146.

47. Nettels, D., Gopich, I.V., Hoffmann, A. & Schuler, B. (2007) Ultrafast dynamics of protein collapse from single-molecule photon statistics. P Natl Acad Sci USA, 104, 2655-2660.

48. Masante, C., Jaubert, C., Palau, W., Plissonneau, J., Besnard, L., Ventura, M. & Di Primo, C. (2015) Mutations of the SL2 dimerization sequence of the hepatitis C genome abrogate viral replication. Cell Mol Life Sci, 72, 3375-3385.

49. Yi, M. & Lemon, S.M. (2003) 3' nontranslated RNA signals required for replication of hepatitis C virus RNA. J Virol, 77, 3557-3568.

50. Bai, Y., Zhou, K.H. & Doudna, J.A. (2013) Hepatitis C virus 3 ' UTR regulates viral translation through direct interactions with the host translation machinery. Nucleic Acids Res, 41, 7861-7874.

51. Romero-Lopez, C., Barroso-DelJesus, A., Garcia-Sacristan, A., Briones, C. & Berzal-Herranz, A. (2014) End-to-end crosstalk within the hepatitis C virus genome mediates the conformational switch of the 3 ' X-tail region. Nucleic Acids Res, 42, 567-582.

52. Shi, G., Ando, T., Suzuki, R., Matsuda, M., Nakashima, K., Ito, M., Omatsu, T., Oba, M., Ochiai, H., Kato, T. *et al.* (2016) Involvement of the 3' Untranslated Region in Encapsidation of the Hepatitis C Virus. PLoS Pathog, 12, e1005441.

53. Shetty, S., Kim, S., Shimakami, T., Lemon, S.M. & Mihailescu, M.R. (2010) Hepatitis C virus genomic RNA dimerization is mediated via a kissing complex intermediate. RNA, 16, 913-925.

54. Sharma, K.K., de Rocquigny, H., Darlix, J.L., Lavergne, J.P., Penin, F., Lessinger, J.M. & Mely, Y. (2012) Analysis of the RNA chaperoning activity of the hepatitis C virus core protein on the conserved 3'X region of the viral genome. Nucleic Acids Res, 40, 2540-2553.

55. Cantero-Camacho, A. & Gallego, J. (2015) The conserved 3'X terminal domain of hepatitis C virus genomic RNA forms a two-stem structure that promotes viral RNA dimerization. Nucleic Acids Res, 43, 8529-8539.

56. Fuxreiter, M. & Tompa, P. (2012) Fuzzy complexes: a more stochastic view of protein function. Adv Exp Med Biol, 725, 1-14.

57. Tompa, P. & Fuxreiter, M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem Sci, 33, 2-8.

58. Aguado-Llera, D., Goormaghtigh, E., de Geest, N., Quan, X.J., Prieto, A., Hassan, B.A., Gomez, J. & Neira, J.L. (2010) The basic helix-loop-helix region of human neurogenin 1 is a monomeric natively unfolded protein which forms a "fuzzy" complex upon DNA binding. Biochemistry, 49, 1577-1589.

59. Kunkel, M. & Watowich, S.J. (2002) Conformational changes accompanying self-assembly of the hepatitis C virus core protein. Virology, 294, 239-245.

60. Wahl, M., Rahn, H.J., Rohlicke, T., Kell, G., Nettels, D., Hillger, F., Schuler, B. & Erdmann, R. (2008) Scalable time-correlated photon counting system with multiple independent input channels. The Review of scientific instruments, 79, 123113.

61. D'Souza, V. & Summers, M.F. (2004) Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus. Nature, 431, 586-590.

62. D'Souza, V. & Summers, M.F. (2005) How retroviruses select their genomes. Nat Rev Microbiol, 3, 643-655.

63. Prasad, B.V. & Schmid, M.F. (2012) Principles of virus structural organization. Adv Exp Med Biol, 726, 17-47.

64. Schmidt, M. (2004) *Polyelectrolytes with defined molecular architecture*. Springer, Berlin ; New York.

65. Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A. *et al.* (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res, 44, D279-285.

66. Kapanidis, A.N., Lee, N.K., Laurence, T.A., Doose, S., Margeat, E. & Weiss, S. (2004) Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. P Natl Acad Sci USA, 101, 8936-8941.

67. Jensen, M.R., Zweckstetter, M., Huang, J.R. & Blackledge, M. (2014) Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. 114, 6632-6660.

68. Schuler, B., Lipman, E.A., Steinbach, P.J., Kumke, M. & Eaton, W.A. (2005) Polyproline and the "spectroscopic ruler" revisited with single-molecule fluorescence. Proceedings of the National Academy of Sciences of the United States of America, 102, 2754-2759.

69. Zhou, H.X. (2004) Polymer models of protein stability, folding, and interactions. Biochemistry, 43, 2141-2154.

70. Cantor, C. & Schimmel, P. (1980) *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules*. First ed. W.H. Freeman and Company, San Fransisco, CA, USA.

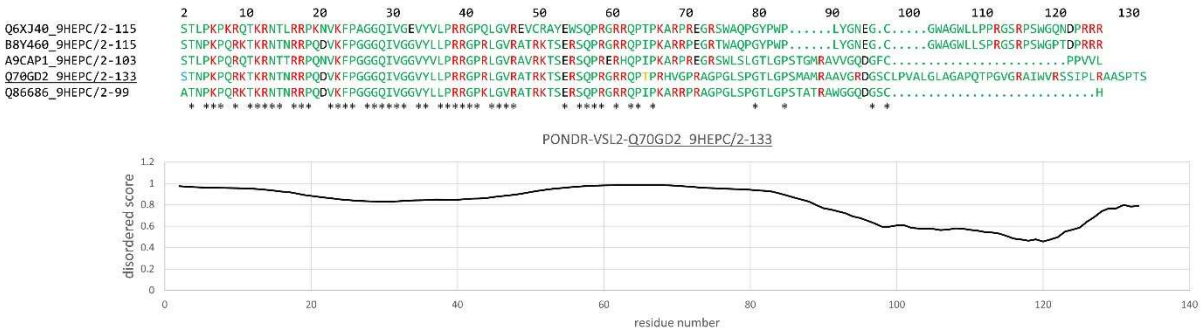
71. Enderlein, J. (2009) In Hinterdorfer, P. and Oijen, A. (eds.), *Handbook of Single-Molecule Biophysics*. Springer US, New York, NY, pp. 243-263.

72. Owczarzy, R., Tataurov, A.V., Wu, Y., Manthey, J.A., McQuisten, K.A., Almabrazi, H.G., Pedersen, K.F., Lin, Y., Garretson, J., McEntaggart, N.O. *et al.* (2008) IDT SciTools: a suite for analysis and design of nucleic acid oligomers. Nucleic acids research, 36, W163-169.

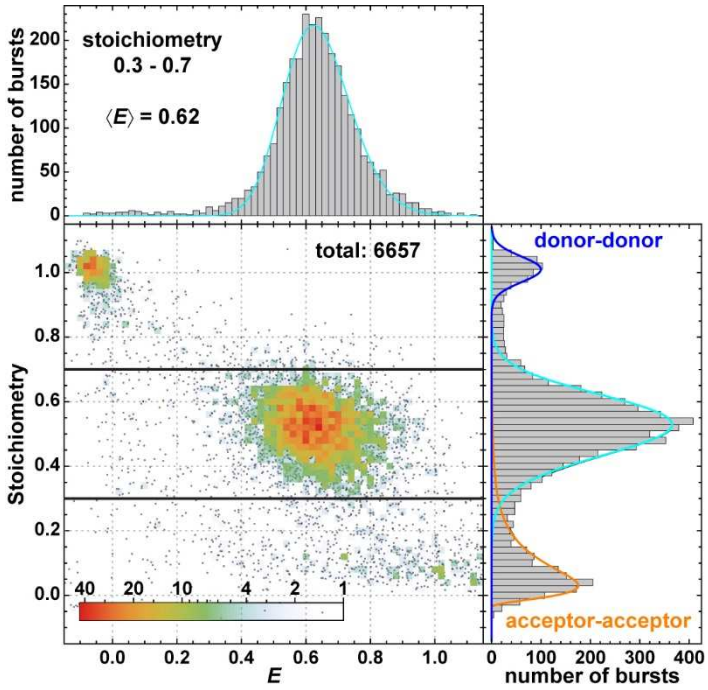
73. Haenni, D., Zosel, F., Reymond, L., Nettels, D. & Schuler, B. (2013) Intramolecular distances and dynamics from the combined photon statistics of single-molecule FRET and photoinduced electron transfer. The journal of physical chemistry. B, 117, 13015-13028.

74. Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K. & Uversky, V.N. (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta, 1804, 996-1010.

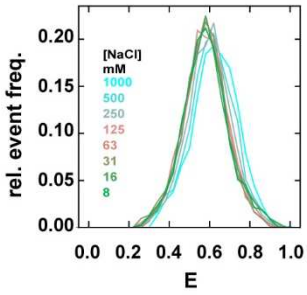
10.0 SUPPLEMENTAL DATA



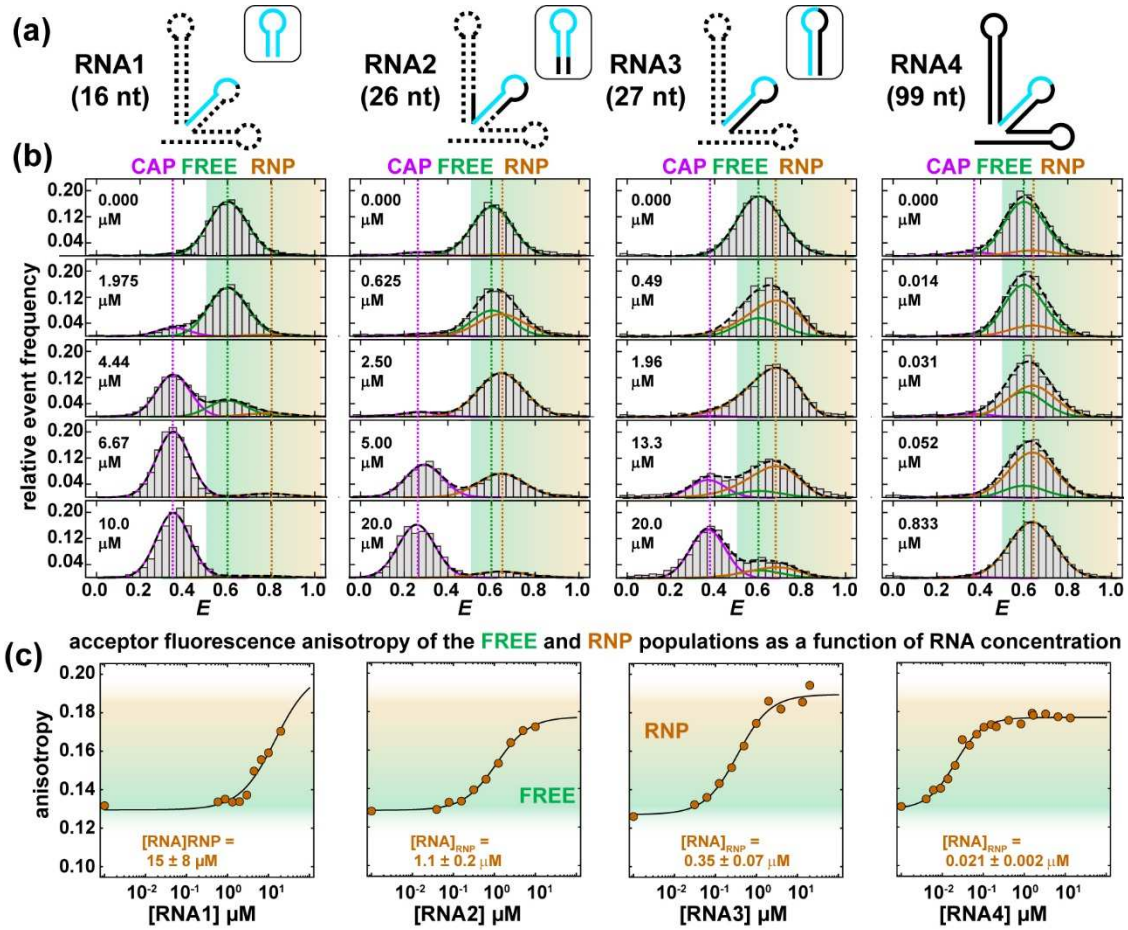
Supplemental Figure 1: Choice of the hepatitis C virus nucleocapsid domain (HCVncd) variant. (top) Multiple sequence alignment for five representative sequences from the HCVncd family (PF01543) in the Pfam database **[65]**. Asterisks (*) indicate complete conservation across all five seed sequences. The sequence Q70GD2 was used in this study because it contains the fewest number of tryptophan residues, which can cause quenching of the FRET dyes and interfere with the quantitative analysis of transfer efficiencies and protein dynamics **[73]**; red font indicates positively charged residues, black font indicates negatively charged residues, blue and orange font indicate the location of the cysteine mutations used for labeling. The sequence separation of the dyes was chosen in order to achieve a mean inter-dye distance, $\langle r_{\text{dye-dye}} \rangle$, that is similar to the Förster radius of this FRET pair, R_0 = 54 Å. As a result, we are able to work in a regime where FRET is most sensitive to distance changes. (bottom) Predicted disorder using the PONDR **[74]** (Predictor of Natural Disordered Regions) predictor VLS2.



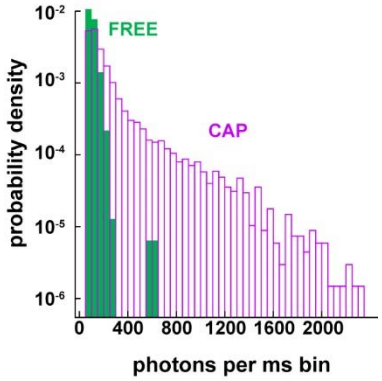
Supplemental Figure 2: Homogeneity of FRET-labeled sample assessed by pulsed interleaved excitation. Each data point in the transfer efficiency vs. stoichiometry plot represents a 1-ms burst event associated with a freely diffusing HCVncd. A projection of the data onto the stoichiometry axis (right) reveals that most bursts arise from FRET-labeled HCVncd molecules containing active donor and acceptor fluorophores. The mean transfer efficiency, ⟨*E*⟩, associated with these molecules is obtained by fitting a histogram of transfer efficiencies from all bursts with a stoichiometry value between 0.3 and 0.7 to a log-normal function (top).



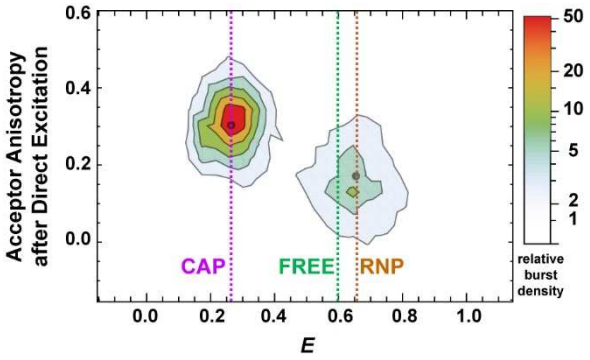
Supplemental Figure 3: Salt concentration has a small effect on the dimensions of HCVncd. (a) *E* histograms for HCVncd at increasing concentrations of NaCl in the presence of 5 mM GdmCl and 50 mM sodium phosphate, pH 7.0.



Supplemental Figure 4: Transition from FREE to RNP probed by subpopulation-specific fluorescence anisotropy. (a) Secondary structure diagrams of the four RNA sequences from the viral genome (Figure 1b) used to induce nucleocapsid formation. Dashed lines correspond to regions of the 3′ X-RNA that were excised in each of the four nucleic acids; the cyan line represents the dimer linkage sequence (DLS). Insets show the predicted secondary structure **[72]** for each of the three truncated sequences. (b) The acceptor fluorescence anisotropy was determined for all bursts within the shaded region to monitor the transition from the FREE conformation to the RNP conformation. (c) RNA binding results in HCVncd molecules with higher acceptor fluorescence anisotropy. Fits are used to determine the transition midpoint, $[RNA]_{RNP}$, associated with RNA binding. Solid lines represent fits to a cooperative binding model. Importantly, the transition midpoints are consistent with the values determined using the transfer efficiency data (Figure 3).



Supplemental Figure 5: The bright fluorescence bursts of the CAP population of HCVncd. Photon count rate probability density histogram, for thousands of 1 ms bursts containing more than 75 photons after donor excitation when accounting for the experimental correction factors (see MATERIALS AND METHODS). In the absence of RNA (i.e., the FREE population, green), nearly all bursts contain 75 - 300 photons. In the presence of 20 μM RNA1 (i.e., the CAP population, purple), bursts are substantially more likely to contains more than 300 photons. These additional photons are probably the result of multiple (~5-10) FRET-labeled capsomeres in each freely diffusing NLP, which contain 100s of individual HCVncd capsomeres.



Supplemental Figure 6: Subpopulation-specific fluorescence anisotropy of HCVncd in different conformations. Acceptor anisotropy vs *E* at 5 μM RNA3, where both the CAP and RNP populations are present. The large anisotropy (~0.33) associated with the CAP conformation indicates orientational restriction of the dyes in the large, slowly tumbling nucleocapsid-like particles comprised of FRET-labeled HCVncd capsomeres. In the RNP population, the average anisotropy is ~0.18, indicating that the fluorophores have more orientational freedom.